DOCUMENT RESUME

ED 450 336 CG 030 793

AUTHOR	Armistead, Leigh; Armistead, Rhonda; Breckheimer, Steve
TITLE	Grading the North Carolina Student Accountability Standards. Student Accountability Standards and High-Stakes Testing in North Carolina: A Position Statement and Supporting Paper.
INSTITUTION	North Carolina School Psychology Association, Charlotte.
PUB DATE	2001-01-00
NOTE	26p.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Academic Achievement; *Academic Standards; Disadvantaged; Elementary Education; Grade 5; Grade 8; *Mathematics Achievement; Professional Associations; *Reading Achievement; School Psychologists; State Curriculum.Guides; *State Standards; *Test Reliability; *Test Validity
IDENTIFIERS	North Carolina; North Carolina End of Grade Testing Program

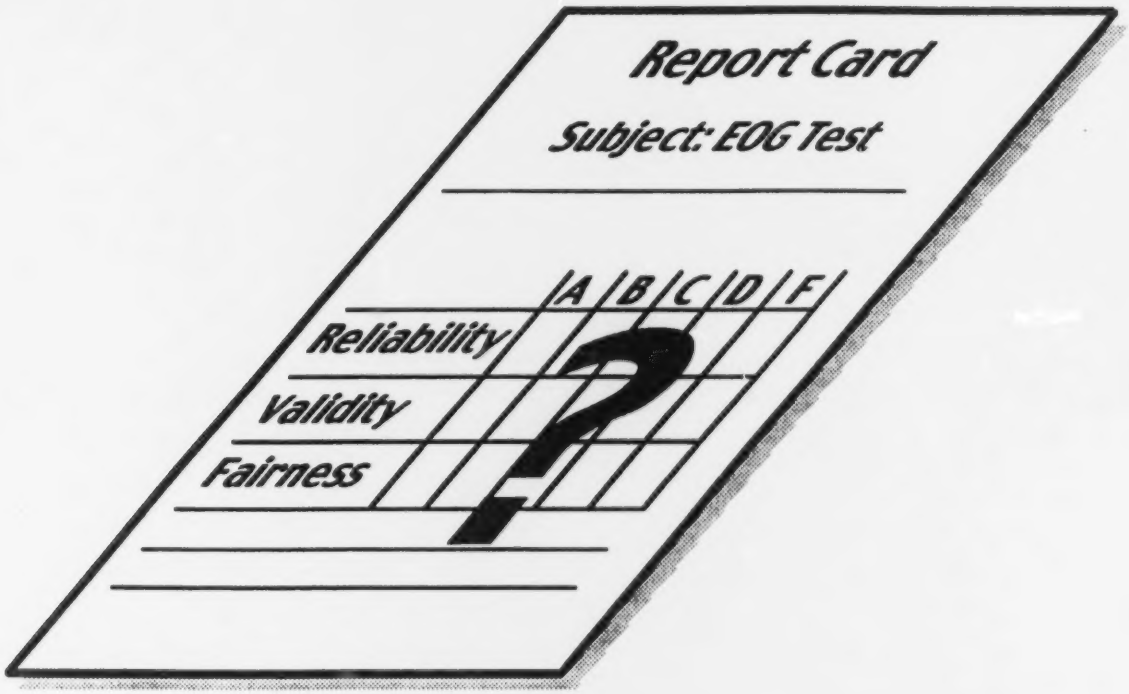
ABSTRACT

School psychologists are in a unique position to add to the discussion about accountability efforts and the effect on students, teachers, and education. At the end of North Carolina's 2000-2001 school year, End-of-Grade (EOG) scores will be used to hold individual students accountable for their own achievement. Fifth graders will be required to score a Level III on both EOG reading comprehension and math in order to be promoted, and next year eight graders will face similar gateways. The North Carolina School Psychology Association contends that use of Student Accountability Standards (SAS) to make major decisions about individual students is not adequately validated and will cause serious harm to the state's most vulnerable students. They question the fairness of EOG test results and the disproportionate impact of certain aspects of the SAS on minority and culturally disadvantaged students, economically disadvantaged students, and students with limited English language. This report points out how EOG test results does not measure up to established standards of reliability, validity, and fairness necessary for making decisions about individual students. Several arguments are presented on the detrimental effects of retention on students. Alternatives to this method are presented that support student learning and help prevent student failure. (Contains 30 references.) (Author/JDM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 450 336

Grading the North Carolina Student Accountability Standards



Student Accountability Standards and High-Stakes Testing in North Carolina: A Position Statement And Supporting Paper

North Carolina School Psychology Association

January 2001

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

L. ARMISTEAD

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

2

CG030793

North Carolina School Psychology Association

Position Statement

Student Accountability Standards And High-Stakes Testing

In May 1995, the North Carolina State Board of Education issued *The New ABCs of Public Education: Accountability, Curriculum Basics and Local Control and Flexibility.* The ABCs included a plan to hold each of the state's schools accountable for the educational growth of groups of students over time. Since then, North Carolina End-of-Grade Test (EOG) scores in reading comprehension and mathematics for grades 3–8 (and writing scores for grades 4 and 7) have been entered into a complex formula to measure and recognize individual school performance and determine financial bonuses for teachers.

At the end of the 2000–2001 school year, however, EOG scores will be used for the first time to hold *individual students* accountable for their own school achievement. Current published materials and comments by state officials have emphasized that single EOG scores will not be the only determinant of promotion; however, confusion about any flexibility in the standards persists.

The North Carolina School Psychology Association (NCSPA) supports high standards for all students. However, supported by an extensive review, we contend that the Student Accountability Standards' (SAS) use of EOG test results to make major decisions about individual students is not adequately validated and will cause serious harm to North Carolina's most vulnerable students. The EOG was not developed for making important decisions about individual students and its use may result in a disregard for additional relevant information from parents, teachers, school staff and the students themselves. In addition, the SAS does not adequately take into account the following:

- The importance of making key, life-changing decisions about students using an array of information, not just test scores.
- The requirement that standardized tests used for making decisions about individual students must meet a higher technical standard than those used for comparing groups of students.
- Extensive research showing that children develop at widely varied times and rates. They learn to walk and talk at different ages and learn academic skills at different rates.
- National standards for the development and use of standardized tests.
- Decades of research showing that retention generally results in no lasting academic benefit, harmful emotional effects, and an increased rate of students' dropping out of school.
- Although retention with extensive remediation has been effective with certain groups of children, promotion with similar remediation is more effective and has fewer negative effects.
- Strong evidence that the Student Accountability Standards will disproportionately affect poor and minority students.
- The current cost of retaining 60,000 students in grades K–12 each year—approximately $360 million—will likely increase as more students are retained.

- Effective alternatives to both retention and social promotion exist.
- The narrowing of the curriculum to the detriment of pupils, teachers and the mission of schools.
- The need for major reform in the way we teach children, organize our schools and fund education in North Carolina.

Therefore, NCSPA's primary recommendation is for North Carolina's State Board of Education to put its implementation of the Student Accountability System on hold while it studies the issues raised in this document. We believe this action is warranted given problems with the SAS and its negative effects on children which are discussed in the background report. We encourage the Board to continue to use the North Carolina End-of-Grade Tests as originally intended—as measures of school improvement at the district and school levels. The background report, of which this is a summary, supports the following additional recommendations:

Effective Practices to Support Student Learning And Prevent Failure

Student Accountability Standards

1. Continue to promote high standards for all students.
2. For individual students, use the EOG scores for screening purposes to determine if students may need additional assistance in those subjects.
3. Change the wording in the Student Accountability Standards to make clear the intended flexibility in the policy and inform stake-holders about the flexible intent of the standards. This could dispel the atmosphere of fear has which developed as a result of ambiguous communication about the standards.
4. Revise the Student Accountability Standards to eliminate the district-level review committees. Instead, require each school to form its own committee to review waiver requests from teachers and parents and make recommendations to the principal regarding promotion and resources needed for the students to be successful. This will ensure that decisions about students will be made, using an array of information, by the people who have worked with and know the students best.
5. Emphasize promotion of students with increased instructional time and special assistance rather than retention. Distribute a summary of current research findings on retention to every school principal and include it in materials provided to any review teams.
6. Modify the SAS policy related to students with limited English proficiency to align it with the research on second language acquisition. Review current research in this area to promote and support effective model programs and develop alternative assessment systems measuring English acquisition.

School Reform

7. Continue to promote class size reduction in grades K–3.
8. Encourage the development of programs that increase parent involvement and create a positive atmosphere for learning. An example of an effective reform effort of this type is the Yale University Child Study Center's Comer Process used in many schools in North Carolina. Showcase these programs at state conferences and in "best practices" publications.
9. Promote the development of broad-based, innovative changes in the schools such as preschool education programs for at-risk children, continuous progress programs in each subject and ungraded classes in grades K–5.
10. Provide leadership to school districts in adopting effective, research-based reading programs which can prevent early failure to acquire basic reading skills.
11. Identify and promote model programs that network school and community resources to address personal and family factors which affect learning.

Testing and Accountability Program

12. Advise school districts that all individual EOG test results should be interpreted with appropriate caution because of the large margin of error in the scores.
13. Develop a statistical reporting system to determine the effects of the EOG testing program. Monitor progress of retained students and determine the relationship between retention and dropping out of school.
14. Continue the development of authentic assessment of student learning instead of relying solely on multiple choice testing.
15. Contract with an independent evaluation team not associated with the development of the testing program to review the program, compare it with the most recent testing standards, and make recommendations for improvement.
16. Do not add field test items to the EOG tests. The tests are lengthy and additional items may change the conditions of the test and invalidate the results. Also set a schedule for completing field tests at the beginning of the school year and stick to it.
17. Require all new and revised tests to be field tested, normed, evaluated, and ready prior to their utilization. This includes the Computerized Adaptive Testing System (CATS).
18. Set standards for appropriate test preparation to increase fairness to all students.

Funding

18. Provide funding for test preparation materials for all schools.
19. Equalize funding across the state so that every child will have the same opportunities and a fairer playing field.
20. Increase funding for intervention efforts with at-risk students. Encourage those efforts in grades K–2 where intervention can have the greatest impact.
21. Provide funding for training of school teams to decrease the number of inappropriate referrals to special education.
22. Fund the development of high quality preschool programs for "at-risk" 4-year-olds.

The Background Report for this Position Statement is available at:
www.ncschoolpsy.org

North Carolina School Psychology Association

Background Report

Student Accountability Standards And High Stakes Testing

The North Carolina School Psychology Association (NCSPA) is a professional organization whose purpose is to serve the educational and mental health needs of students, assist with the development of sound educational practices, and advance the practice of school psychology. School psychologists are in a unique position in education because of their training in both psychology and education. They espouse a scientist-practitioner perspective which will enrich the discussion about the accountability effort and its effects on students, teachers, and education.

In May 1995, the North Carolina State Board of Education issued *The New ABCs of Public Education: Accountability, Curriculum Basics and Local Control and Flexibility.* The ABCs included a plan to hold each of the state's schools accountable for the educational growth of groups of students over time. Since then, North Carolina End-of-Grade Test (EOG) scores in reading comprehension and mathematics for grades 3–8 (and writing scores for grades 4 and 7) have been entered into a complex formula to measure and recognize individual school performance and determine financial bonuses for teachers.

At the end of the 2000-2001 school year, however, EOG scores will be used for the first time to hold *individual* students accountable for their own school achievement. Fifth graders will be required to score a Level III on both the EOG Reading Comprehension and Mathematics exams in order to be promoted to 6th grade. In the 2001-2 school year, 3rd and 8th graders will face similar "gateways." The potential impact of these new promotion standards is now becoming clearer:

The Charlotte Observer has published an analysis of the May 2000 EOG test results and reported that about one third of this state's current 5th graders are at risk of being retained in 2001—more than 30,000 students. (Rothacker and Mellnik, 2000, August 21). Other *Observer* findings:

- At the end of the 99-00 school year, 12,308 fourth grade students in North Carolina were below grade level in reading and math, 15,407 were below grade level in reading only, and 2,858 were below grade level in math only.
- In Mecklenburg County, 2,710 fifth graders are at risk for retention. The Mecklenburg County data also illustrates the disparate impact on minorities. Of African-American 5th graders, 51% are at risk for retention whereas just 15% of white students are at risk.
- At one inner-city Mecklenburg County school which has had the benefit of state intervention teams in the past, 75% of the 5th graders are now at risk for retention.

During the 1999–2000 school year, Wilson County required all students in grades 3–8 to score at Level III or higher to be promoted. In grades 6, 7 and 8, a total of 389 students were retained. At $6000 per student (the average state expenditure), retaining these Wilson County students alone will eventually cost over $2,000,000.

Even before implementation of the SAS, the current statewide annual retention rate is about 5%. More than 60,000 students are retained each year in grades K-12. Using the average state expenditure of $6000 per student, the estimated *current* statewide cost of retentions is $360,000,000 per year. This estimate does not include capital costs for additional schools and classrooms that will be required to give retained students an additional year of education.

Published materials and comments by state officials have emphasized that single EOG scores will not be the only determinant of promotion. However, confusion about any intended flexibility in the standards persists, possibly because the standards emphasize that students must score a Level III or higher to be considered on grade level. Students will be permitted to take the EOG a second or third time and a Personal Education Plan specifying focused intervention will be provided for each student scoring at Level I or II on the EOG. An appeal process will also be available and principals will continue to have the final authority to promote or retain students. However, given the large number of students at risk, and the current SAS policy, it seems likely that EOG scores will be the primary criteria used to decide promotion or retention for many. Therefore, it seems prudent to examine the appropriateness of using the North Carolina End-of-Grade Test for such high stakes decisions as promotion and retention of individual students. This paper will focus on three aspects of the North Carolina Student Accountability System:

- The EOG and High-Stakes Decisions About Children
- The Effectiveness of Retention
- Fairness of the SAS for Subgroups of Children

The paper will conclude with a review of the many factors involved in children's learning and recommendations for changes in the SAS.

The EOG and High-Stakes Decisions About Children

The use of standardized tests such as the EOG for educational decision making seems quite popular with the majority of Americans. A recent poll by Public Agenda, a nonprofit, nonpartisan public policy research organization, found that 71% of parents support testing during elementary school to help identify struggling students. Seventy-five percent agreed that, "students pay more attention and study harder if they know they must pass a test to get promoted or to graduate." (Public Agenda, 2000)

Dr. Aaron M. Pallas, professor of sociology and education at Teachers College, Columbia University, and co-author of a report on high-stakes testing for the Civil Rights Project at Harvard University has sought to explain this popularity. In a recent issue of the *Harvard Education Letter*, Dr. Pallas states,

> Most standardized tests are viewed by the public at large as objective, which means several things: there are right and wrong answers to the test questions; unlike grades, which are awarded at the whim of a teacher, standardized tests are standardized—scores don't depend on who is performing the assessment; tests yield numerical scores, which are precise measures of performance; and, like a laboratory measurement, test scores are reliable. Testing experts acknowledge that some of these assumptions are questionable. Test construction is a social and political process, and we cannot afford to lose sight of that fact. (Sadowski, 2000).

Several resources are available to help evaluate the objectivity and appropriateness of standardized tests for specific purposes. To ensure that tests are used appropriately, the US Congress directed the National Academy of Sciences through its National Research Council (NRC) to study the issue of high stakes testing and make recommendations. Those recommendations were published in 1999 in the report *High Stakes: Testing for Tracking, Promotion and Graduation* (Heubert and Hauser, 1999).

The American Education Research Association (AERA) has revised its *Standards for Educational and Psychological Tests*. These standards represent a consensus of several professional organizations concerning sound and appropriate test use in education

(AERA, 1999). Recently, AERA also issued a position statement on high-stakes testing (AERA, 2000).

Finally, the U.S. Department of Education, Office for Civil Rights (OCR) has recently released a draft version *The Use of Tests When Making High-Stakes Decisions for Students: A Resource Guide for Educators and Policy makers.* (OCR, 2000).

These reports make many recommendations about using standardized tests in an appropriate and legal manner. The following standards are relevant to a review of the technical adequacy of the EOG and will be discussed in the following sections:

Reliability: Tests are not perfect; scores vary. It must be established that a student's scores are reliable enough to support the intended interpretations of those scores. A test cannot be considered valid unless its results are reliable.

Validity: The important thing about a test is not its validity in general, but its validity for a specific purpose. A test can be valid for one purpose but not for another; the validity of each separate use of a test must be evaluated separately. The validity of cut scores and achievement levels must also be established.

Fairness: Besides the technical attributes of reliability and validity, tests should embody social values of equity and justice and, for example, not systematically underestimate the achievements of a particular group.

Reliability of the EOG

The technical manual for the EOG, *North Carolina End of Grade Tests, Technical Report #1* (Sanford, 1996), includes this definition: "Reliability refers to the consistency of scores obtained by the same person when examined by the same test on different occasions or with different sets of equivalent items. If any use is to be made of the information from a test, then it is desirable that the test results be reliable." (p. 45) Test reliability is usually expressed by reliability coefficients which range from .00 (no reliability) to 1.00 (perfect reliability).

The EOG *Technical Report* presents the results of just two reliability studies. The first study examined internal consistency reliability or the extent to which items on the test all measure the same characteristic. The reliability coefficients from this study of the 1993 administration of the test were all .90 or higher. For a test that measures a single subject such as mathematics, we should expect high internal consistency. The EOG appears to be reliable with regard to internal consistency.

However, other forms of reliability are probably more relevant when evaluating an individual student's test results. Test-retest reliability coefficients, for example, provide an indication of how stable test results are over time. The second study described in the *Technical Report* looked at a combination of test-retest and alternate form reliabilities. A second version of the 7th grade reading test was given to three classes (70 students) in one North Carolina school district a week after they took the first version. The reliability estimate obtained was .86. The manual does not indicate whether the mathematics test was administered and, if it was, what reliability estimate was obtained.

The EOG *Technical Report* states that, "If decisions about individuals are to be made on the basis of the test data (for example, placement or instructional program decisions), then it is desirable that the test results be reliable and exhibit a reliability coefficient of at least .85." (p. 45). It appears that, for the 7th grade reading test at least, the EOG meets this criterion. Because younger children have had less experience with standardized tests, it seems likely that third graders' test score reliability quotients would be lower than those obtained for 7th graders. The Office for Civil Rights *Testing Guide* recommends that, "reliability data should be presented as soon as feasible for each major population for whom the test is recommended." Data for grade levels other than 7th grade are not included in the *Technical Report.*

The *Technical Report's* contention that a reliability coefficient of .85 is sufficient could be questioned. Salvia and Ysseldyke (1991), well-regarded specialists in the area of

standardized testing, recommend a reliability coefficient of at least .90 when test scores are used for important individual decisions.

Another way of looking at the reliability of a test is to look instead at its unreliability. Unreliability is indexed by the standard error of measurement of a test. This index can be used to define a range of uncertainty for a test score which is similar to the familiar margin of error (plus or minus a certain number of points) which is reported for public opinion poll results. The EOG *Technical Report* reports that the standard error for most students is 2 to 3 points. While this error may sound trivial, the following scenario provides a different perspective on the importance of the standard error:

> Imagine a 5th grade student with an EOG score of 140 in reading. She needs to score 149 to be on Level III and be promoted to 6th grade. Her standard error of measurement is 4 points. Since the state's scoring program makes an allowance for one standard error, our student would actually need a lower score than 149—probably just 146. Because of the unreliability of test scores, test developers specify a range of scores which they believe contains a child's true score. With the standard error of 4 points, we can be 90% confident that her true score is between 132 and 148. This range includes the 146 score she probably needs for promotion to 6th grade.

Another way to gain perspective on the unreliability of the EOG is to compare its standard error with expected growth from one year to the next. The state-wide average growth in EOG reading scores between 5th and 6th grades is just 3 points. This is actually less than the 4-point standard error of measurement for the 5th grade Level II reader cited in the previous scenario.

The EOG *Technical Report* does not provide adequate evidence of the reliability of the EOG for high-stakes decision making. However, a DPI official has been quoted as saying that a student's taking the EOG the second or third time (prior to retention) will improve the reliability of the student's score. No studies to support this claim have been offered and it seems likely that some students may not take advantage of retesting since the present policy makes it optional. A further problem with this retesting-to-provide-reliability rationale is that students who score at Level III will, in most cases, be promoted automatically on the basis of a single, and perhaps unreliable, score. No additional testing will be conducted to determine the reliability of their scores.

Validity of the EOG

As noted previously, tests can be regarded as valid only for specific purposes. The *Technical Report* for the EOG states that the EOG was developed for two purposes:

> "to provide accurate measurement of individual student skills and knowledge specified in the North Carolina Standard Course of Study," and
>
> "to provide accurate measurement of the knowledge and skills attained by groups of students for school, school system, and state accountability" (p. 1).

The manual does not mention the EOG's use for accountability of *individual* students except for this comment on page three, "For individual student accountability, the grade eight end-of-grade tests are used as a way for students to demonstrate that they have the knowledge and skills necessary to meet the reading and mathematics competency requirement for high school graduation."

Although, the original purpose of the EOG was not for deciding the fate of individual students, it still could be used in that way if it met generally accepted technical standards, especially with regard to validity. The NRC report states, "It should be clear that what needs to be validated is not the test in general or in the abstract, but rather each inference that is made from the test scores and each specific use to which the test is put. Although there is a natural tendency to use existing tests for new and different purposes, each new purpose must be validated in its own right." (Heubert and Hauser, 1999).

For a standardized test, validity essentially means: a) the test measures what it purports to measure (and only what it is supposed to measure) and, b) the conclusions to be drawn from the test are meaningful. Because tests are used for many different purposes, there is no single type of validity evidence that is appropriate for all intentions. The EOG *Technical Report* discusses three types of validity evidence for the EOG:

1. *Content validity* refers to whether test content includes an appropriate sample of the knowledge and skills that are the goals of instruction (Sattler, 1992). The EOG *Technical Report* documents that adequate content validity was built into the EOG as it was developed. All items are described as aligned with the North Carolina Standard Course of Study. Items were written and reviewed by North Carolina classroom teachers in a process that is well documented in the manual. It appears that the EOG measures what it is supposed to measure.

2. *Criterion-related validity* refers to relationships between test scores and an outcome—a rating, a classification, or another test score (Sattler, 1992). There are two kinds of criterion-related validity. Concurrent validity refers to a relationship with some measure of rating currently available. An example would be a test's agreement with current student grades. Predictive validity refers to a test's relationship with some future performance. An example would be a test's ability to predict future student grades.

The EOG *Technical Report* includes information for just one type of criterion-related validity—the relationship of EOG scores to teacher judgments about current achievement levels. During the field testing of the EOG in 1992, teachers were asked to rate each student who took the test into one of these categories:

- Level I (*Fails to achieve at a basic level*): Students performing at this level do not have sufficient mastery of knowledge and skills in this subject area to be successful at the next grade level.
- Level II (*Achieves at a basic level*): Students performing at this level demonstrate inconsistent mastery of knowledge and skills that are fundamental in this subject area and that are minimally sufficient to be successful at the next grade level.
- Level III (*Achieves at a proficient level*): Students performing at this level consistently demonstrate mastery of grade level subject matter and skills and are well prepared for the next grade level.
- Level IV (*Achieves at an advanced level*): Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient at grade level work, or
- Not a clear example of any one of these achievement levels.

These descriptions are almost identical to those used today to describe student achievement levels except that the brief descriptors shown in italics above have now been removed. It might be considered ironic that students who were described by their teachers in 1992 as achieving, "at a basic level" subsequently would now be classified as candidates for retention.

Over 160,000 students were included in the 1992 field test. Teachers categorized about 95% of their students into one of the four achievement levels. Across most grade levels and in both reading and math, about 40% of students were rated as performing in Levels I and II and about 60% were rated as performing in Levels III and IV at the time they took the EOG.

The EOG manual cites the relationship between teacher judgments of student's achievement levels and their concurrent EOG scores as evidence of the test's criterion-related validity. However, no correlation coefficients are provided as is usual when reporting such results. Instead the manual includes the two diagrams which are reproduced below as evidence of criterion-related validity. Each vertical line shows the range of scores earned by the middle two thirds of each achievement group within each grade level. For example, the middle two thirds of the 5th graders rated by their teachers

as Level II students scored between approximately 140 and 154. Of course, about 33% of the 5th graders must have scored above or below this range.

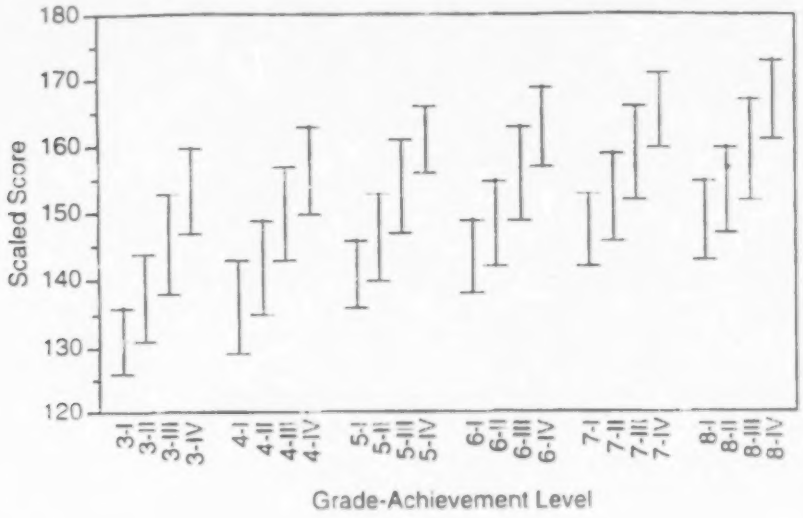


Figure 23. The relationship between teacher judgments of student achievement and scores on the North Carolina End-of-Grade Test of Reading Comprehension field test (May 1992).

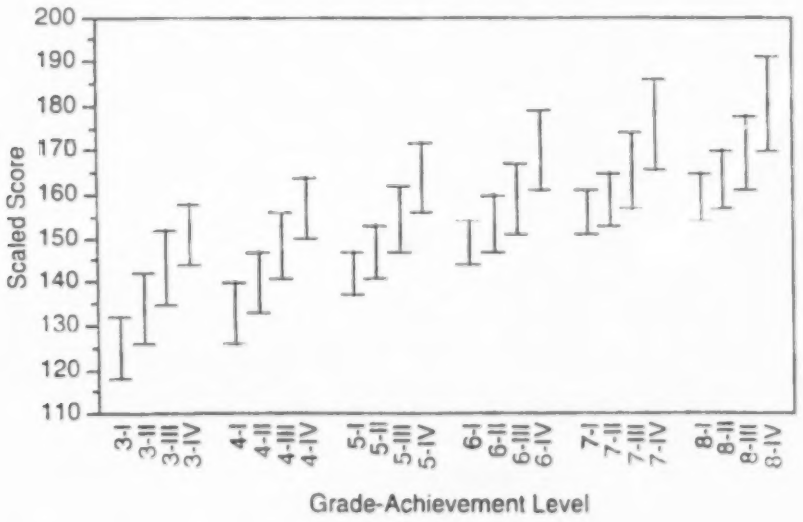


Figure 24. The relationship between teacher judgments of student achievement and scores on the North Carolina End of Grade Test of Mathematics field test (May 1992).

The EOG *Technical Report* cites these tables as evidence of validity stating, "As expected, the scaled scores increase over the achievement levels, and also across grades. Students rated by their teachers as high achievers (Level IV) scored high on the tests.

while students who were rated low by teachers scored low on the test (Level I)." (Sanford, 1996, p. 51)

This type of validity evidence might be considered adequate for a test intended to measure progress of large groups of students from one year to the next, perhaps to assess the performance of school districts and individual schools. However, at the individual student level, it is not adequate. Although the test and teacher ratings were in agreement for the majority of students, thousands of students were incorrectly rated by their teachers as compared to their test scores *or* inaccurately assessed by the test as compared to their teachers' ratings. Nevertheless, the teacher ratings from the field test were used to establish the cut points now being used by schools to determine promotion or retention.

The second type of criterion-related validity mentioned previously is predictive validity—the ability of a test to predict a future outcome such as performance in the next grade level. The National Research Council report emphasizes that this type of validity evidence is especially important, "when using test scores for selection, placement, certification of competence, program evaluation, and other kinds of accountability." (Heubert and Hauser, 1999, p. 76)

The EOG *Technical Report*, however, does not present any predictive validity data. There is no information provided about how the approximately 64,000 students classified by their teachers as Level I or II (that is, not ready or only minimally ready for the next grade level) actually performed in the next grade level. Nevertheless, the cut scores for promotion and retention established by those teacher ratings will be used by schools to promote and retain students. The Office for Civil Rights draft report advises states and school districts that cut scores used for high stakes decision making must be validated for that purpose. Validity of the EOG cut scores which designate whether students are below or on grade level is not adequately addressed in the *Technical Report*.

3. The third type of validity to be considered, *construct validity*, refers to the extent to which a test measures a particular construct or trait such as mathematics achievement or reading comprehension. As evidence of construct validity, the EOG *Technical Report* includes correlations between the EOG and selected tests that could be considered to measure the same constructs. Correlations are presented with the North Carolina Open-Ended Tests. They ranged from the mid .50s for reading to the mid .60s for math. A correlation of .73 was found between the 8th grade EOG reading and math results and the 9th grade end-of-course tests in English I and Algebra I. Portions of the Iowa Tests of Basic Skills were administered to a sample of 5th and 8th graders in 1993. Correlations ranged from .76 to .84. Finally, portions of the 1992 National Assessment of Educational Progress (NAEP) math test were administered to a sample of 8th graders. A correlation of .70 was reported.

These correlations can best be interpreted as showing that when groups of students take similar tests with similar content, they tend to get similar rankings of scores. Unfortunately, it is not possible to evaluate the construct validity of the EOG on the basis of the results presented in the *Technical Report* because they are incomplete. The Manual does not include, for example, correlations with the NAEP Reading test or any data for grade levels other than 8th grade on the NAEP.

Fairness of the EOG

Fairness of a test relates to its, "comparable validity, that is whether it provides comparably valid scores across individuals, groups, and settings." (Heubert and Hauser, 1999, p. 78). With regard to fairness, the Office for Civil Rights report contends,

> Demonstrating fairness in the validation of test score inferences focuses primarily on making sure the scores reflect the same intended knowledge and skills for all students taking the test. For the most part, this means that the test should minimize the measurement of material that is extraneous to the intended constructs and which confounds the ability of the test to accurately measure the constructs

that it intends to measure. Rather a test score should accurately reflect how well each student has mastered the intended constructs. The score should not be significantly impacted by construct-irrelevant influences. (OCR, 2000, p. 28–9)

As previously noted, 1999 4th grade EOG scores suggest that a higher percentage of African-American students are at risk of failing 5th grade than of white students. The OCR *Testing Guide* points out that such disparities are not sufficient to establish a violation of civil rights laws. However, the *Guide* contends that such disparities create the need for further examination of the educational practices that have caused the disparities thus ensuring nondiscriminatory decision making.

For example, standardized tests such as the EOG have sometimes been criticized for including racially biased questions. This aspect of fairness was addressed during the development of the EOG. Each item was statistically checked for possible gender bias and racial bias (blacks and whites only). Items that were flagged by this process were examined by a group of individuals representing various minority groups and by curriculum specialists. Items consistently identified as biased were removed from the pool of test items.

However, it is possible that the design of the EOG itself could contribute to score disparities among various groups because the test does not just measure whether or not a student is on grade level, as is commonly believed. It also *ranks* students in comparison with other students. To understand this, one must consider how the EOG was designed.

There are two general approaches to standardized test design: *norm-referenced testing* and *criterion-referenced testing*. Norm-referenced testing evaluates a student's performance compared to the performance of others on the same test. When you consider a score report that shows a child performed at the 50th percentile, you know that she scored higher than 50% of the children who took the test. In other words, children are ranked from the lowest to highest scores. In contrast, criterion-referenced testing is used to measure a student's status with regard to an established level of performance. It measures degree of mastery. A test which shows a child got 75% of the answers correct is criterion-referenced. There is no comparison with other children. Theoretically, it is possible for every student to get every answer correct or to get 100% mastery.

The EOG, however, attempts to provide both criterion- and norm-referenced information about students. It provides the Level I, II, III and IV scores which purport to indicate mastery of grade level material. (As noted previously, there are serious questions about the reliability and validity of these level scores.) In addition, the EOG gives scaled scores and percentile scores that rank each student in comparison with other students. To do so, the EOG was constructed to provide a range of scores or ranking of students. The EOG *Technical Report* discloses that each question on the EOG is not just aligned with the curriculum but is also classified along two dimensions: *difficulty level* and *thinking skills* level.

Difficulty Level. With regard to difficulty level, the EOG was constructed so that:

25% of items are easy (definition: can be answered by 70% of examinees),

50% of items are at the medium level (definition: can be answered by 50-60% of students), and

25% of items are at the difficult level (definition: can be answered by only 20 or 30% of students)

This difficulty level dimension means that no matter how good the instructional program and student effort level become, it is not theoretically possible for every student to answer every question correctly as they would on a criterion-referenced test.

Thinking Skills. The second dimension, the thinking skills level, refers to the, "cognitive skills that a student must employ to solve the problem." (Sanford, 1996, p.10) A

philosophy or framework called *Dimensions of Thinking* (Marzano, et. al, 1988) was used to develop questions for the EOG. It is a complex framework which includes:

1. Content Area Knowledge
2. Metacognition (ability to think about your own thinking)
3. Critical and Creative Thinking
4. Core Thinking Skills, or "building blocks" of thinking including
   a. focusing
   b. information-gathering
   c. remembering
   d. organizing
   e. analyzing
   f. generating
   g. integrating
   h. evaluating
5. Thinking Processes or relatively complex sequences of thinking skills:
   a. concept formation
   b. principle formation
   c. comprehending
   d. problem solving
   e. decision-making
   f. research
   g. composing
   h. oral discourse

As previously discussed, EOG items can be regarded as closely aligned with the North Carolina curriculum. However, the thinking skills framework discussed here receives scant mention in the NC curriculum. The introduction to the NC Standard Course of Study provides a much briefer description of the *Dimensions of Thinking* framework. Also included are what are called, "guiding assumptions for a thinking framework for North Carolina's public schools." Here are examples of these assumptions:

"All students can become better thinkers."

"Thinking is improved when the learner takes control of his/her thinking processes and skills."

"The teaching of thinking should be deliberate and explicit...." (DPI, 1999, p. xi)

In the sections which follow the introduction, only occasional, and somewhat vague, references to these thinking skills can be found. For example, 6th grade Competency Goal 5: "The learner will respond to various literary genres using interpretive and evaluative processes." (DPI, 1999, p. 61).

Even the answer format of the EOG was selected with the thinking skills framework in mind. The "best answer" format used was chosen because, "this format is well suited for testing a student's ability to evaluate (Marzano's highest thinking skill level)." Each item was evaluated so that incorrect answers, "should appear plausible for someone who has not achieved mastery of the representative objective." (Sanford, 1996, p. 28)

The EOG's combination of difficulty level and thinking skills framework spreads out the scores and results in a normal distribution of scores, much like the distribution curve of scores from an intelligence test. As do intelligence tests, the EOG stresses the ability to apply information in new and different ways rather than just mastery of learned information.

The fairness of using a test that measures many aspects of academic aptitude or cognitive ability to determine promotion or retention of individual students is

questionable. It is likely that low-ability students, who can acquire core academic skills, will not be able to demonstrate their mastery of those skills on the EOG. And it is those "core academic competencies" which the EOG was originally intended to measure (Public School Law 115C-174.11(c)).

National testing guidelines suggest that it *is* appropriate for DPI to encourage the teaching of high-level thinking skills and even include assessment of those skills in a test intended to assess schools. However, it seems unfair to make important life decisions about students based on a test that partly measures thinking skills. As the National Research Council points out, a test can be valid for one purpose and not another. "Tests that are valid for influencing classroom practice, 'leading' the curriculum, or holding schools accountable are not appropriate for making high-stakes decisions about individual student mastery unless the curriculum, the teaching and the test(s) are aligned." (Heubert and Hauser, 1999, p. 3)

A final issue of fairness of the EOG relates to current test administration practices. The ABCs program includes financial incentives for teachers in schools which achieve specified goals. This has resulted in a complex set of procedures to maintain test security and prevent what are termed "administration irregularities" by teachers who administer tests. Teachers and proctors, for example, are prevented from telling students about technical mistakes such as marking answers in the wrong section of the test booklet or misaligning their answers and question numbers. The North Carolina Association of Educators (NCAE) has recognized this fairness issue and has called for removing, "from our classrooms a 'Gotcha' mentality that prevents teachers from helping children take tests correctly...let us at least ensure that students' scores are determined on the basis of knowledge rather than technicalities." (NCAE, 2000, p.1).

The first section of this background paper has shown that the North Carolina EOG test does not measure up to established standards of reliability, validity and fairness necessary for making important decisions about individual children. However, even if it were shown to meet these standards, it should not be used to make bad decisions. The next section will review the evidence regarding retention and show that it is usually harmful to students.

The Effectiveness of Retention

One rationale for the SAS is to end the practice of social promotion in North Carolina's schools. The implication is that social promotion is rampant and retention is rare. However, the rate of retention in grade in NC has actually increased from 3.2% of all public school students in 1992–93 to 5.0% during the 1997–99 period (North Carolina Department of Public Instruction, 2000 a). This amounts to retaining about 60,000 students in grades K–12 each year (not including 20,000 students who were promoted after attending summer school). It is clear that, in North Carolina, social promotion is not the norm and retention is not rare.

The practice of retaining students in a grade has been extensively studied over several decades and the preponderance of results show that retained students do worse academically than comparable students who are promoted. Retention has also been shown to have negative effects on personal adjustment, attitudes towards school and school drop out rates. (Dawson and Rafoth, 1998). A sample of the research findings on retention:

- Some groups of students are more likely to be retained than others. Those at highest risk for retention tend to: be Black or Hispanic, have late birthdays (e.g., August, September, October), have developmental delays and/or attention problems, live in poverty, live in a single-parent household, have parents with low educational attainment, or have changed schools frequently (National Association of School Psychologists, 1998).

- Early retentions are not better than later ones. There is no evidence of positive effects on school achievement or personal adjustment by such practices as delayed entry into school, kindergarten retention, or transitional classes.
- Reading is the primary academic problem for which students are retained.
- Initial achievement gains may occur during the first year of retention, but a consistent finding across many research studies is that such achievement gains decline within 2–3 years. Retained children either do no better or perform more poorly that similar groups of promoted children. This is true whether children are compared to same–age or same–grade students who were promoted. One of the reasons that teachers often underestimate the negative effects of retention is that these effects may not show up until the student is in another grade or school.
- Children who are developmentally delayed are most likely to be harmed by retention. Particularly at the first grade level, large percentages of retained children are either subsequently retained again or are placed in special education.
- Retention is associated with significant increases in behavior problems, with problems becoming more pronounced as children reach adolescence.
- Students who are retained drop out of school at a much higher rate than promoted students even controlling for prior achievement, grades and attendance (Roderick, 1995). This finding is true whether the retention occurs early or late in their school career. For students who have been retained twice, the likelihood of dropping out increases by 90% (Task Force on Education of Young Adolescents, 1989).
- A recent national longitudinal study shows that the use of high stakes 8th grade tests is associated with sharply higher drop-out rates, especially for students at schools serving mainly low SES students (Reardon, 1996).
- Asked to rate stressful experiences, a group of students rated only blindness and death of a parent as more stressful than being retained in school (Byrnes and Yamamoto, 1984).

Some have argued that retention research has not looked at what has been called "retention with remediation." They correctly point out that a few studies have provided some support for retention. However, Holmes (1990) points out that these studies are similar in that they occurred in suburban settings, and included few, if any, disadvantaged students. Most retained students had average IQs and near-average reading skills. Retained students were not recycled through the standard curriculum but were placed in special classes with low teacher/pupil ratios and given considerable extra help. It should be noted that most of these successful retention studies did not provide remediation for the at-risk *promoted* children with whom they compared the retained children. In those that did, promoted at-risk children with extra help did better than retained children with extra help.

William Romey has pointed out that it's ironic that, "Retaining a child who hasn't passed a certain level at the end of June isn't really retention at all. It is moving the child clear back to the beginning of the year he or she has failed rather than working with the individual child at his or her actual achievement level." (Romey, 2000, p. 632). Romey suggests that children do not need to repeat an entire grade when they are missing part of the material—they just need to practice some of the material longer.

Despite the consistent research findings that retaining students does not improve long-term achievement and will actually increase the chances of dropping out of school, the SAS emphasizes retention as a way of helping students.

Fairness of the Student Accountability System for Subgroups of Children

Previous sections have briefly noted the disproportionate impact of certain aspects of the SAS on minority and culturally disadvantaged children. It is also important to consider how various subgroups of children are likely to be affected.

Disadvantaged Children

The 2000 EOG results indicate an average of 24% of all students are below grade level on the composite EOG scores (reading, math and writing) in grades 3 through 8. The breakdown of these results by race shows 18% of White students, 34% of Hispanic and American Indian students, and 40% of African-American students scoring below grade level (North Carolina Department of Public Instruction, 2000b). Fifty percent of Limited English Proficient students scored below grade level on the composite score. Students who are on free lunch or whose parents have less than a high school diploma are also more likely to score below grade level.

While the achievement gap between disadvantaged and more affluent students is widely acknowledged, the causes of the differences is multifactored. Socioeconomic disparities apparently play a major role since educational achievement correlates more strongly with economic status than with any other single variable. Other social factors such as unstable families, poor parenting skills, teen pregnancy, drugs, crime, poor role models, and lack of parent involvement are considered by some to be significant barriers to academic success. (Singham, 1998).

English Language Learners

North Carolina's growing population of students with limited English language proficiency are also likely to be negatively affected by the SAS. Despite exemptions and waivers for students who are learning English as a second language, the SAS appears to disregard the timeline for acquiring a second language. For many children, only two years of instruction may be needed to acquire basic conversational skills. Cummins (1984), however, has shown that five to seven years are needed for students to acquire what is known as *cognitive/academic language proficiency* (CALP) which is necessary for understanding English during context-reduced academic situations such as reading the passages on the EOG reading test.

Children with Disabilities

According to Federal law, students with various disabilities must also participate in state testing programs to the "extent possible." Starting in 2001, only the most severely disabled students will be exempted from the EOG testing program. Most students with disabilities will be required to take the regular EOG test or, if they are at least two years below grade level and meet other requirements, the Computerized Adaptive Testing System (CATS). In a November 9, 2000 *Assessment Brief*, DPI announced that the CATS system will initially include EOG test questions from the final semester of 2nd grade through 10th grade. With the CATS version of the EOG, multiple choice test questions will be presented on a computer screen. The computer will adjust the difficulty level of subsequent questions up or down depending on the student's accuracy. (How this will work for a third grader with a first grade reading level is unclear.)

The CATS is being developed to allow what's known as *out-of-level testing*, that is, permitting a student with a disability who is below grade level to take a below-grade test. Students who are below grade level because of limited English proficiency or because they have low intellectual ability or because they have an economically or socially disadvantaged background will apparently not be permitted to participate in out-of-level testing.

Children in Grades K–2

It might appear that the EOG and SAS can only affect students in grades 3–8 and that children in grades K–2 are not affected. It is true that children in these grades cannot be assessed with group standardized tests. The North Carolina Legislature banned such testing in 1987. The North Carolina Association for the Education of Young Children (NCAEYC), the Atlantic Center for Research in Education (ACRE), and NCSPA strongly supported the legislation's adoption. The ban was based on an awareness that, "group standardized testing will not improve individual achievement or educational

standards, but will undermine the solid gains made in the education of young children in North Carolina in recent years. The consequences will be to distort the curriculum, divert resources away from our successful early childhood program, and cause harm to many children, particularly children with special needs." (NCAEYC, 1998a)

It does appear, however, that K–2 students are being affected by the SAS through developmentally inappropriate instructional practices, downward pressure on the curriculum and early identification of students who might not "pass the test." In some districts, K-2 students are being categorized as "on grade level" or "below grade level" depending on their performance on informal tests. Although this could be a positive consequence if more resources were directed toward students at risk of academic failure, it could also result in K–2 students being retained earlier in an attempt to "prevent" later retentions.

Narrowing of the Curriculum

Disadvantaged children are affected more than other children by what's been called the "narrowing" of the curriculum. Preparing students for the SAS has apparently required teachers to focus on basic reading, math and writing skills. A wide range of strategies have been used to accomplish this, but the most common one seems to be to increase the time for teaching the "core" subjects and to reduce the time allocated for other subjects such as science, social studies, physical education, music and art. While this may result in an increase in test scores in the core subjects, an unintended effect could be to produce students who are less knowledgeable about the physical and political world and who are less physically fit. Middle class parents may be able to compensate for this narrowing of the curriculum; disadvantaged parents may not.

A second strategy for improving test scores in reading, math and writing is the use of test preparation materials and practice tests. The result has been to significantly increase the amount of time devoted to student preparation for the EOG. Eighty percent of teachers in one survey stated that their students spent more than 20% of their instructional time practicing for the test (Jones, et.al., 1999). Since test preparation materials (not practice tests) are purchased locally, students in more wealthy school systems would seem to have a significant advantage over students in less wealthy districts.

Proponents of the SAS assert that holding school districts and teachers accountable with the ABCs is not enough—that children need to be held accountable also. This has led to one of the more insidious effects of the SAS on children: its oversimplification of the complex web of relationships involving instruction, curriculum, and learner characteristics. An implication of the SAS is that children, with their teachers' help, simply have to try harder—try harder to get on grade level, try harder to achieve level III, and try harder to be promoted to the next grade. Although effort is important, children and their learning are much more complicated than that. To support this contention, in the next section we present 14 psychological principles that pertain to children and their learning process.

Learner-Centered Psychological Principles

The American Psychological Association's Board of Educational Affairs has published a summary of psychological principles related to the learner and learning. These principles are based on more than a century of research on learning and teaching and are widely utilized in effective schools (American Psychological Association, 1987).

These principles emphasize the active and thoughtful nature of learning and learners. They focus on psychological factors that are primarily internal to and under the control of the learner rather than conditioned habits or physiological factors. However, the principles also attempt to acknowledge external environment or contextual factors that interact with these internal factors.

The principles are intended to deal holistically with learners in the context of real-world learning situations. Thus, they are best understood as an organized set of principles; no principle should be viewed in isolation. The 14 principles are divided into four factors which influence learners and learning: *cognitive and metacognitive, motivational and affective, developmental and social, and individual difference* factors. Finally, the principles are intended to apply to all learners—from children, to teachers, to administrators, to parents, and to community members involved in our educational system.

Cognitive and Metacognitive Factors

Teachers are presented with students every day who embody a range of individual abilities and levels of prior knowledge. The most effective learning process is one which helps the learner construct meaning from information and experiences by making the learning active, goal directed and relevant to each student. Teachers play a major interactive role with both the learner and the learning environment. Effective teaching builds links between existing knowledge bases and new information. To obtain this result, strategies such as concept mapping and thematic organization or categorizing have been shown to be effective with learners of varying abilities.

As acknowledged by the North Carolina Department of Instruction's Standard Course of Study, helping students to develop strategic thinking is important to achieving complex learning goals. Successful learners create and use a repertoire of thinking and reasoning strategies to solve problems and learn new concepts. Thus, educators can enhance learning outcomes by assisting students to develop, apply and assess strategic learning skills. In turn, students will continue to expand their repertoire of strategies by reflecting upon the methods which work well, by receiving guided instruction and feedback and by observing additional models. This process for developing critical thinking is largely a cumulative one and may culminate for some students only after years of effective schooling.

Learning does not occur in a vacuum. Cultural or group influences on students can impact many educationally relevant variables, such as motivation, orientation toward learning and ways of thinking. Instructional practices and technologies must appropriate for learners' level of prior knowledge, cognitive abilities, and their learning and thinking strategies.

Motivational and Affective Factors

What and how much is learned is influenced by the learner's motivation. Motivation to learn, in turn, is influenced by the individual's emotional states, beliefs, interests and habits of thinking. A student's internal thoughts, beliefs and expectations for success or failure can enhance or interfere with his or her quality of thinking and information processing. Students' beliefs about themselves as learners have a marked influence on motivation. In turn, motivational and emotional factors also influence both the quality of thinking as well as an individual's motivation to learn. Positive emotions, such as curiosity, generally enhance motivation and facilitate learning and performance. However, intense negative emotions (e.g., anxiety, panic, rage, insecurity) and related thoughts (e.g., worrying about competence, ruminating about failure, fearing punishment, ridicule, or stigmatizing consequences) generally detract from motivation, interfere with learning, and contribute to low performance. Intrinsic motivation is more likely to be achieved when students perceive a task as interesting, personally relevant and meaningful, appropriate for his or her abilities, and on which they believe they can succeed. One of the most consistent and robust findings in the area of motivation is the importance of self-efficacy to performance. One key finding related to self-efficacy is that learners must attribute success to effort and strategies. Emphasizing a student's improvement over time, rather than comparing a student's performance to other students, is likely to increase the student's self-efficacy for learning.

Effort is another major indicator of motivation to learn. Learning complex knowledge and skills demands that learners invest high levels of energy and focused effort, along with persistence over time. Teachers must be concerned with facilitating motivation by using strategies that increase effort and a commitment to learning. Effective strategies include learning activities with high task value, practices that enhance positive emotions and intrinsic motivation to learn and methods that increase learner's perceptions that a task in interesting and personally relevant.

Developmental and Social Factors

Individuals learn best when material is appropriate to their developmental level and is presented in an enjoyable and interesting way. As humans, our individual development varies across intellectual, social, emotional, and physical domains and thus achievement in different instructional domains may also vary. The cognitive, emotional, and social development of individual learners and how they benefit from life experiences are affected by home, prior schooling, cultural, and community factors. Awareness and understanding of developmental differences among children with and without physical, intellectual or emotional disabilities, is necessary to create optimal learning contexts.

Learning can be enhanced when the learner has an opportunity to interact and to collaborate with others on instructional tasks. Learning settings that allow for social interactions, and that respect diversity, encourage flexible thinking and social competence. In interactive and collaborative instructional contexts, individuals have an opportunity for perspective taking and reflective thinking that may lead to higher levels of cognitive, social, and moral development, as well as self-esteem. Quality personal relationships that provide stability, trust, and caring can increase learners' sense of belonging, self-respect and self-acceptance, and provide a positive climate for learning. Positive learning climates can also help to establish the context for healthier levels of thinking, feeling, and behaving.

Individual Differences

Individuals are born with and develop their own capabilities and talents. In addition, they have acquired their own preferences for how they like to learn and the pace at which they learn. However, these preferences are not always useful in helping learners reach their learning goals. Educators need to help students examine their learning preferences and expand or modify them, if necessary. The interaction between learner differences and curricular and environmental conditions is another key factor affecting learning outcomes.

The same basic principles of learning, motivation, and effective instruction apply to all learners. However, language, ethnicity, race, beliefs, and socio-economic status all can influence learning. Careful attention to these factors in the instructional setting enhances the possibilities for designing and implementing appropriate learning environments. When learners perceive that their individual differences in abilities, backgrounds, cultures, and experiences are valued, respected, and accommodated in learning tasks and contexts, levels of motivation and achievement are enhanced.

Assessment provides important information to both the learner and teacher at all stages of the learning process. Effective learning takes place when learners feel challenged to work towards appropriately high goals; therefore, appraisal of the learner's cognitive strengths and weaknesses, as well as current knowledge and skills, is important for the selection of instructional materials of an optimal degree of difficulty. Ongoing assessment of the learner's understanding of the curricular material can provide valuable feedback to both learners and teachers about progress toward the learning goals. Self-assessments of learning progress can also improve students self appraisal skills and enhance motivation and self-directed learning.

Conclusions and Recommendations

The North Carolina School Psychology Association contends that the North Carolina Student Accountability Standards' use of EOG test scores to make major decisions about individual students is not adequately validated and will cause serious harm to North Carolina's most vulnerable students. The EOG was not developed for making important decisions about individual students and its use may result in a disregard for additional relevant information from parents, teachers, school staff and the students themselves. In addition, the SAS does not adequately take into account the following:

- The importance of making key, life-changing decisions about students using an array of information, not just test scores.
- The requirement that standardized tests used for making decisions about individual students must meet a higher technical standard than those used for comparing groups of students.
- Extensive research showing that children develop at widely varied times and rates. They learn to walk and talk at different ages and learn academic skills at different rates.
- National standards for the development and use of standardized tests.
- Decades of research showing that retention generally results in no lasting academic benefit, harmful emotional effects, and an increased rate of students' dropping out of school.
- Although retention with extensive remediation has been effective with certain groups of children, promotion with similar remediation is more effective and has fewer negative effects.
- Strong evidence that the Student Accountability Standards will disproportionately affect poor and minority students.
- The current cost of retaining 60,000 students in grades K-12 each year—approximately $360 million—will likely increase as more students are retained.
- Effective alternatives to both retention and social promotion exist.
- The narrowing of the curriculum to the detriment of pupils, teachers and the mission of schools.
- The need for major reform in the way we teach children, organize our schools and fund education in North Carolina.

Therefore, NCSPA's primary recommendation is for North Carolina's State Board of Education to put its implementation of the Student Accountability System on hold while it studies the issues raised in this document. We believe this action is warranted given problems with the SAS and its negative effects on children which are discussed in this background report. We encourage the Board to continue to use the North Carolina End-of-Grade Tests as originally intended—as measures of school improvement at the district and school levels. We believe that this report supports the following recommended alternatives to, or modifications of, the SAS.

Alternatives to the Student Accountability System

North Carolina's children are quite diverse—ethnically, socially, economically, and developmentally. They vary in the age at which they enter school, and they develop at differing, individual rates throughout their school experience. North Carolina has an established curriculum for each grade level. Children move through the curriculum in unique and individual ways. Given true criterion-referenced assessment tools, teachers could assess their students' progress through the curriculum, and determine which areas need further instruction and when a particular student is ready to move on. Teachers are in a better position to do this kind of assessment than any single EOG-style standardized test. Teachers know that "mastery" means consistent demonstration of a

skill rather than demonstrating or not demonstrating a skill on a single test on a single day.

Retention and social promotion are both failed practices. Research on child development suggests that interventions should be provided for children even before they enter school. Unfortunately, most children in North Carolina do not have access to the kinds of high quality early childhood environments that would prepare them for academic excellence. Many child care programs lack an appropriate curriculum and qualified teachers. Improving the quality of early childhood education would be a better intervention than subsequent retention, and it is one recommended by the National Association for the Education of Young Children (NAEYC) and the International Reading Association in a joint position statement *Learning to Read and Write...* (NAEYC, 1998) b). Expanding North Carolina's Smart Start initiative would also provide better preschool programs in North Carolina.

William Romey has written, "As long as we continue to accept that schools must be organized into archaic grade levels, the problem of promotion will plague us." (2000, p. 632) Given the diversity cited above, it is time to seriously consider nongraded, multi-age approaches to organizing classrooms in the elementary school years. This would allow children to learn foundation skills in an early childhood setting, with each child evaluated individually using portfolio assessments and promoted to the next level of learning as he or she is ready. This would also challenge quicker students who have already mastered material at their grade level.

Effective Practices to Support Student Learning And Prevent Failure

Student Accountability Standards

1. Continue to promote high standards for all students.
2. For individual students, use the EOG scores for screening purposes to determine if students may need additional assistance in those subjects.
3. Change the wording in the Student Accountability Standards to make clear the intended flexibility in the policy and inform stake-holders about the flexible intent of the standards. This could dispel the atmosphere of fear has which developed as a result of ambiguous communication about the standards.
4. Revise the Student Accountability Standards to eliminate the district-level review committees. Instead, require each school to form its own committee to review waiver requests from teachers and parents and make recommendations to the principal regarding promotion and resources needed for the students to be successful. This will ensure that decisions about students will be made, using an array of information, by the people who have worked with and know the students best.
5. Emphasize promotion of students with increased instructional time and special assistance rather than retention. Distribute a summary of current research findings on retention to every school principal and include it in materials provided to any review teams.
6. Modify the SAS policy related to students with limited English proficiency to align it with the research on second language acquisition. Review current research in this area to promote and support effective model programs and develop alternative assessment systems measuring English acquisition.

School Reform

7. Continue to promote class size reduction in grades K–3.
8. Encourage the development of programs that increase parent involvement and create a positive atmosphere for learning. An example of an effective reform effort of this type is the Yale University Child Study Center's Comer Process used in many schools in North Carolina. Showcase these programs at state conferences and in "best practices" publications.

9. Promote the development of broad-based, innovative changes in the schools such as preschool education programs for at-risk children, continuous progress programs in each subject and ungraded classes in grades K–5.
10 Provide leadership to school districts in adopting effective, research-based reading programs which can prevent early failure to acquire basic reading skills.
11. Identify and promote model programs that network school and community resources to address personal and family factors which affect learning.

Testing and Accountability Program

12. Advise school districts that all individual EOG test results should be interpreted with appropriate caution because of the large margin of error in the scores.
13. Develop a statistical reporting system to determine the effects of the EOG testing program. Monitor progress of retained students and determine the relationship between retention and dropping out of school.
14. Continue the development of authentic assessment of student learning instead of relying solely on multiple choice testing.
15. Contract with an independent evaluation team not associated with the development of the testing program to review the program, compare it with the most recent testing standards, and make recommendations for improvement.
16. Do not add field test items to the EOG tests. The tests are lengthy and additional items may change the conditions of the test and invalidate the results. Also set a schedule for completing field tests at the beginning of the school year and stick to it.
17. Require all new and revised tests to be field tested, normed, evaluated, and ready prior to their utilization. This includes the Computerized Adaptive Testing System (CATS).
18. Set standards for appropriate test preparation to increase fairness to all students.

Funding

18. Provide funding for test preparation materials for all schools.
19. Equalize funding across the state so that every child will have the same opportunities and a fairer playing field.
20. Increase funding for intervention efforts with at-risk students. Encourage those efforts in grades K–2 where intervention can have the greatest impact.
21. Provide funding for training of school teams to decrease the number of inappropriate referrals to special education.
22. Fund the development of high quality preschool programs for "at-risk" 4-year-olds.

References

American Educational Research Association (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

American Educational Research Association (2000). *AERA position statement concerning high-stakes testing in PreK-12 education.* [On-line]. Available: www.aera.net/about/policy/stakes.htm.

American Psychological Association (1997). *Learner-Centered psychological principles: A framework for school redesign and reform.* Revision prepared by a Work Group of the American Psychological Association's Board of Educational Affairs (BEA), November 1997.

Byrbes, D. and Yamomto, K. (1984). *Grade repetition: Views of parents, teachers, and principals.* Logan,UT: Utah State School of Education.

Cummins, J. (1984) *Bilingualism and special education: issues in assessment and pedagogy.* San Diego, CA: College-Hill.

Dawson, M. M. & Rafoth, M. A. (1991). Why student retention doesn't work. *Streamlined Seminar, 9,* 3.

Heubert, J. P. & Hauser, R. M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation.* Washington: National Academy Press. [Also on-line document]. Available: www.nap.edu/books/0309062802/html/ index.html

Holmes, C. T. (1990). Grade level retention effects: A meta-analysis of research studies. In L. A. Shepard & M. L. Smith (Eds.) *Flunking grades: Research and policies on retention.* New York: Farmer.

Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., Yarbough, T., & Davis, M. (1999) The impact of high stakes testing on teachers and students in North Carolina. *Phi Delta Kappan, 81,* 3.

Linn, Robert L. (2000, March). Assessments and Accountability. *ER Online* [On-line serial], *29*(2). Available: www.aera.net/pubs/er/arts/29-02/linn01.htm.

Marzano, R. J., Brandt, R. S., Hughes, C. S., Jones, B. F., Presseisen, B. Z., Rankin, S. C., & Suhor, C. (1988.) *Dimensions of thinking: A framework for curriculum and instruction.* Alexandria, VA: Association for Supervision and Curriculum Development.

National Association for the Education of Young Children (1998 a) *Position paper on standardized testing of young children in North Carolina.* Raleigh, NC: Author.

National Association for the Education of Young Children (1998 b). *Learning to read and write: Developmentally appropriate practices for young children.* [On-line]. Available: www.naeyc.org/resources/position_statements/ psread0.htm.

National Association of School Psychologists (1998). *Position statement: Student grade retention and social promotion.* Bethesda, MD: Author. (available at www.naspweb.org).

North Carolina Association of Educators (2000, September). Does the ABCs program pass or fail? *News Bulletin, 31,* 2.

North Carolina Department of Public Instruction. (1999). *North Carolina standard course of study.* Raleigh, NC: Author.

North Carolina Department of Public Instruction. (2000a) *North Carolina public schools statistical profile.* Raleigh, NC: Author.

North Carolina Department of Public Instruction. (2000b) *A report card for the abcs of public education, vol II, 1999-00.* Raleigh, NC: Author.

Public Agenda (2000, October 5). *Survey finds little sign of backlash against academic standards or standardized tests.* [On-line] Available: www.publicagenda.org.

Reardon, S. (1996) *Eighth-grade minimum competency testing and early high school dropout patterns.* Paper presented at the Annual Meeting of the American Educational Research Association. New York, NY.

Roderick, M. (1995). Grade retention and school dropping out: Policy debate and research questions. *Research Bulletin 15.* Center for Evaluation, Development and Research, Phi Delta Kappa.

Romey, W. (2000). A note on social promotion. *Phi Delta Kappan, 81,* 8.

Rotthacker, J. W. & Mellnik, T. (2000, August 21). Test rules get tough on school promotions. *The Charlotte Observer,* pp. A1, A6.

Sadowski, M. (2000, November/December). Are high-stakes tests worth the wager? *Harvard Education Letter,* [On-line serial]. Available: http://edletter.org/past/issues/2000-so/tests.html.

Salvia, J. & Ysseldyke, J. (1991) *Assessment in special and remedial education* (4th ed.). Boston: Houghton Mifflin

Sanford, E. E. (1996). *North Carolina End-of-Grade Tests (Technical Report #1).* Raleigh, NC: Department of Public Instruction.

Sattler, J. (1992). *Assessment of Children.* La Mesa, CA: Jerome M. Sattler.

Singham, M. (1998). The canary in the mine: The achievement gap between black and white students. *Phi Delta Kappan, 79,* 9.

Task Force on Education of Young Adolescents (1989). *Turning points: Preparing American youth for the 21st century.* Washington, DC: Carnegie Council on Adolescent Development.

U.S. Department of Education, Office for Civil Rights (2000). *The use of tests when making high-stakes decisions for students: A resource guide for educators and policy makers.* [On-line]. Available: www.ed.gov/offices/ OCR/testing/Testing Resource.pdf.

Student Accountability System Workgroup

Steve Breckheimer, Ed.S., NCSP
Private Practice

Leigh Armistead, Ed.D., NCSP
Charlotte-Mecklenburg Schools

Rhonda Armistead, M.S., NCSP
Charlotte-Mecklenburg Schools

Lisa Murr, M.S., NCSP
Chapel Hill-Carboro City Schools

Lynne Myers, Ph.D.
Wake County Schools

Alice Wellborn, M.A., NCSP
Transylvania County Schools

North Carolina School Psychology Association
Board of Directors
2001

President: Rhonda Armistead, Charlotte-Mecklenburg Schools
President-Elect: Mark Pisano, Fort Bragg Schools
Secretary: Amy Wichman, Charlotte-Mecklenburg Schools
Treasurer: Nick Myers, Durham Public Schools

South Piedmont Representatives:

Bill Coram, Charlotte-Mecklenburg Schools
Linda Haigh, Charlotte-Mecklenburg Schools

North Piedmont Representatives:

Carol Vatz, Wake County Schools
Diane Kelly, Durham Public Schools

Coastal Representatives:

Sallie Moore, Wayne County Schools
Steven Hardy-Braz, Private Practice

Mountain Representative:

Kathleen Robertson, Wilkes County Schools

Delegate, National Association of School Psychologists:

Leigh Armistead, Charlotte-Mecklenburg Schools

North Carolina School Psychology Association
PO Box 16694
Charlotte, NC 28297-6694

Internet Website
www.ncschoolpsy.org